LETTER

doi:10.1038/nature10074

Evolution and metabolic significance of the urea cycle in photosynthetic diatoms

Andrew E. Allen[1,2], Christopher L. Dupont[1], Miroslav Oborník[3], Aleš Horák[3], Adriano Nunes-Nesi[4]†, John P. McCrow[1], Hong Zheng[1], Daniel A. Johnson[1], Hanhua Hu[2]†, Alisdair R. Fernie[4] & Chris Bowler[2]

Diatoms dominate the biomass of phytoplankton in nutrient-rich conditions and form the basis of some of the world's most productive marine food webs[1–4]. The diatom nuclear genome contains genes with bacterial and plastid origins as well as genes of the secondary endosymbiotic host (the exosymbiont[5])[1,6–10], yet little is known about the relative contribution of each gene group to diatom metabolism. Here we show that the exosymbiont-derived ornithine-urea cycle, which is similar to that of metazoans but is absent in green algae and plants, facilitates rapid recovery from prolonged nitrogen limitation. RNA-interference-mediated knockdown of a mitochondrial carbamoyl phosphate synthase impairs the response of nitrogen-limited diatoms to nitrogen addition. Metabolomic analyses indicate that intermediates in the ornithine-urea cycle are particularly depleted and that both the tricarboxylic acid cycle and the glutamine synthetase/glutamate synthase cycles are linked directly with the ornithine-urea cycle. Several other depleted metabolites are generated from ornithine-urea cycle intermediates by the products of genes laterally acquired from bacteria. This metabolic coupling of bacterial- and exosymbiont-derived proteins seems to be fundamental to diatom physiology because the compounds affected include the major diatom osmolyte proline[12] and the precursors for long-chain polyamines required for silica precipitation during cell wall formation[11]. So far, the ornithine-urea cycle is only known for its essential role in the removal of fixed nitrogen in metazoans. In diatoms, this cycle serves as a distribution and repackaging hub for inorganic carbon and nitrogen and contributes significantly to the metabolic response of diatoms to episodic nitrogen availability. The diatom ornithine-urea cycle therefore represents a key pathway for anaplerotic carbon fixation into nitrogenous compounds that are essential for diatom growth and for the contribution of diatoms to marine productivity.

An ornithine-urea cycle (OUC) driven by mitochondrial carbamoyl phosphate synthase (CPS) was long assumed to have originated in metazoans[13–16] and is known to have facilitated key physiological and life-history adaptations in vertebrates[13–15,17,18]. Genome sequence data for marine diatoms, however, suggested a much earlier origin for the OUC[6,7,19]. Extended phylogenetic analysis and estimates of the timing of CPS diversification confirm that stramenopiles and haptophytes, which diverged much earlier than metazoans, also contain OUC-type CPS (novel CPS-a; Fig. 1). Additionally, it is evident that the *CPS* gene passed through two duplication events. The first duplication (1,181–1,500 million years ago; Fig. 1) pre-dates the emergence of metazoans in the eukaryotic tree (Fig. 1 and Supplementary Fig. 1). The second duplication (1,126–1,402 million years ago; Fig. 1) occurred before the diversification between metazoans, stramenopiles and haptophytes of OUC-type CPS (Fig. 1, clade A). Therefore, this duplication probably preceded the secondary endosymbiotic event(s) and associated plastid acquisitions in the latter two lineages. The pattern and timing of *CPS* duplication and divergence indicate that the secondary endosymbiotic host cell of stramenopile and haptophyte algae shared a common ancestry with metazoans at the point of key metabolic innovations that are absent in the Plantae (Archaeplastida) lineage.

In addition to the expansion of this OUC CPS clade to include basal metazoans such as the sea anemone (*Nematostella vectensis*), two exclusively unicellular eukaryotic clades emerged. Novel CPS-a (Fig. 1) is related to mitochondrial OUC-type CPS enzymes which use ammonium (unCPS) and glutamine (ugCPS), and is found in Stramenopila and Haptophyta. Microscopy and measurements of protein activity in the pennate diatom *Phaeodactylum tricornutum* (Supplementary Figs 2 and 3 and Supplementary Table 1) support a mitochondrial localization for this gene product and indicate specificity for NH_4^+ rather than a preference for glutamine, as is the case in early diverging metazoans[18]. This indicates that the ancestral mitochondrial CPS was dependent on NH_4^+ rather than glutamine[14].

The second new type of CPS, pgCPS2, thus named because it probably uses glutamine in pyrimidine synthesis, is found in stramenopiles, haptophytes and alveolates. This enzyme seems to be the result of the second *CPS* duplication (1,126–1,402 million years ago), which involved replacement of ancestral pgCPS1 with an enzyme related to mitochondrial unCPS. Although pgCPS2 is associated by phylogeny with OUC-related CPS, genes in the pgCPS2 clade do not encode a mitochondrial targeting signal and the proteins encoded are probably glutamine-dependent cytosolic enzymes that perform the first committed step of pyrimidine synthesis[15]. This is consistent with the finding that diatoms synthesize nucleotides in the cytosol and subsequently import them into plastids[20].

To evaluate the role of the diatom OUC, we investigated the timing and overall pattern of gene expression as well as the protein levels of unCPS and other OUC gene products in response to the addition of different nitrogen substrates to nitrogen-depleted *P. tricornutum* cultures. Genes involved in nitrate assimilation were used as a contrast. After nitrogen addition, these sets of genes showed unique but coordinated patterns of expression. In short (24 h) and long (120 h) nitrogen-recovery experiments (Fig. 2 and Supplementary Figs 4–6), genes involved in nitrate assimilation showed a strong upregulation that occurred specifically in response to nitrate and was followed by a quick decline, whereas OUC gene transcription occurred immediately after all nitrogen additions and remained elevated for longer than the nitrate-assimilation transcripts. Core OUC genes responded immediately (1–5 h), whereas downstream OUC-related genes responded on longer timescales (24–120 h), especially in response to NO_3^-. This may be due to the longer period of time required for nitrogen, and associated carbon, from NO_3^- assimilation to cycle through cellular pathways to the OUC. Furthermore, there were similar acute responses in the protein levels of mitochondrial unCPS to all nitrogen substrates

[1]J. Craig Venter Institute, San Diego, California 92121, USA. [2]CNRS UMR8197 INSERM U1024, Environmental and Evolutionary Genomics Section, Institute of Biology, Ecole Normale Supérieure, 46 rue d'Ulm, 75005 Paris, France. [3]Biology Centre ASCR, Institute of Parasitology and University of South Bohemia, Faculty of Science, Branišovská 31, 370 05 České Budějovice, Czech Republic. [4]Max-Planck-Insitut für Molekulare Pfanzenphysiologie, 14476 Potsdam-Golm, Germany. †Present addresses: Max Planck partner group Departamento de Biologia Vegetal, Universidade Federal de Viçosa, 36570-000 Viçosa, MG, Brasil (A.N-N.); Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China (H.H.).

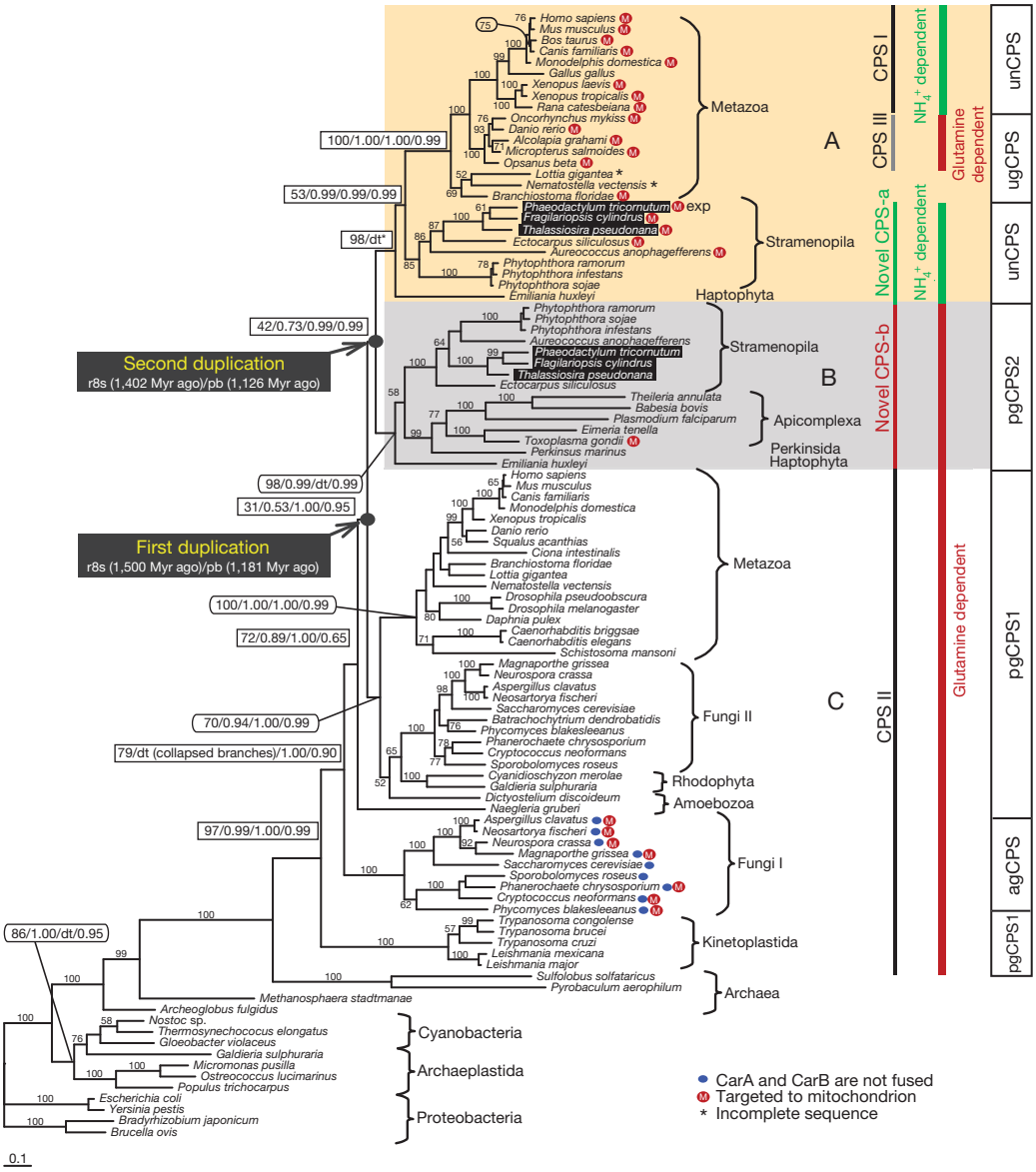

Figure 1 | Carbamoyl phosphate synthase phylogeny and divergence timing. Randomized accelerated maximum likelihood (RaXML with LG model) phylogenetic tree, inferred from amino acid sequences of carbamoyl phosphate synthases (CPS). Numbers above branches indicate RaXML bootstraps/PhyloBayes posterior probabilities/MrBayes posterior probabilities/PhyML. dt, different topology for particular method; exp, targeting was experimentally confirmed in this work. Sequences not composed of a fusion of CPS subunits CarA+B are marked. Abbreviations preceding 'CPS' indicate pathway (a, arginine synthesis; p, pyrimidine synthesis; u, urea cycle) and substrate (g, glutamine; n, ammonium). ugCPS and unCPS (clade A), pgCPS2 (clade B), and pgCPS1 (clade C) are indicated. CPS I, II and III designations have been used in the literature previously. The CPS nomenclature introduced here reflects proposed evolutionary chronology, substrate and pathway. The estimated times of both gene duplications leading to the appearance of the urea cycle are indicated: Myr, million years; Rates (r8S) 1.71 and Phylobayes (pb) 3.2 were used to estimate divergence times. See Supplementary Fig. 1 and Supplementary Tables 3–5. Scale bar indicates 0.1 amino acid substitutions per position.

and of nitrate reductase to nitrate (Supplementary Fig. 4). Urease protein levels, however, were consistently elevated with little variation in response to the source of nitrogen or the timing of the response to nitrogen addition (Supplementary Fig. 4), implying that there is near-constitutive intracellular production and turnover of urea regardless of the nitrogen substrate. Across 16 different conditions[21], expression of OUC genes was elevated in response to conditions favourable for protein synthesis, such as the onset of light, elevated CO_2 and urea as the sole source of nitrogen (Supplementary Fig. 7). This implicates the diatom OUC in anabolic metabolism, in contrast to animal systems where the OUC typically has a catabolic role[22,23].

To gain further insight into the function of the diatom OUC, RNA interference (RNAi) techniques[24] were used to generate *P. tricornutum* cell lines with reduced levels of unCPS. unCPS protein content was reduced in six lines relative to wild type (Supplementary Fig. 8a). These lines also showed a 15–30% reduction in steady-state growth rate when growing on NO_3^- or urea as the sole nitrogen source (Supplementary Fig. 8b). Additionally, when permitted to enter stationary phase, the *unCPS* RNAi lines required a longer period than wild-type cells to recover and resume growth upon transfer to fresh media, suggesting that impairment of mitochondrial unCPS significantly hampers rapid recovery from nitrogen limitation (Supplementary Fig. 8c). Within 48 h of exposure to fresh nitrogen, wild-type cells achieved maximum photosynthetic yield (variable fluorescence/maximum fluorescence, Fv/Fm), maximum growth rates and significant consumption of inorganic carbon. In contrast, the *unCPS* RNAi lines exhibited significant delays (Supplementary Fig. 9). These lags in exponential growth and photosynthetic yield resemble the physiological profile of algal cells treated with protein synthesis inhibitors during recovery from nitrogen limitation[25], indicating that mitochondrial CPS and the OUC

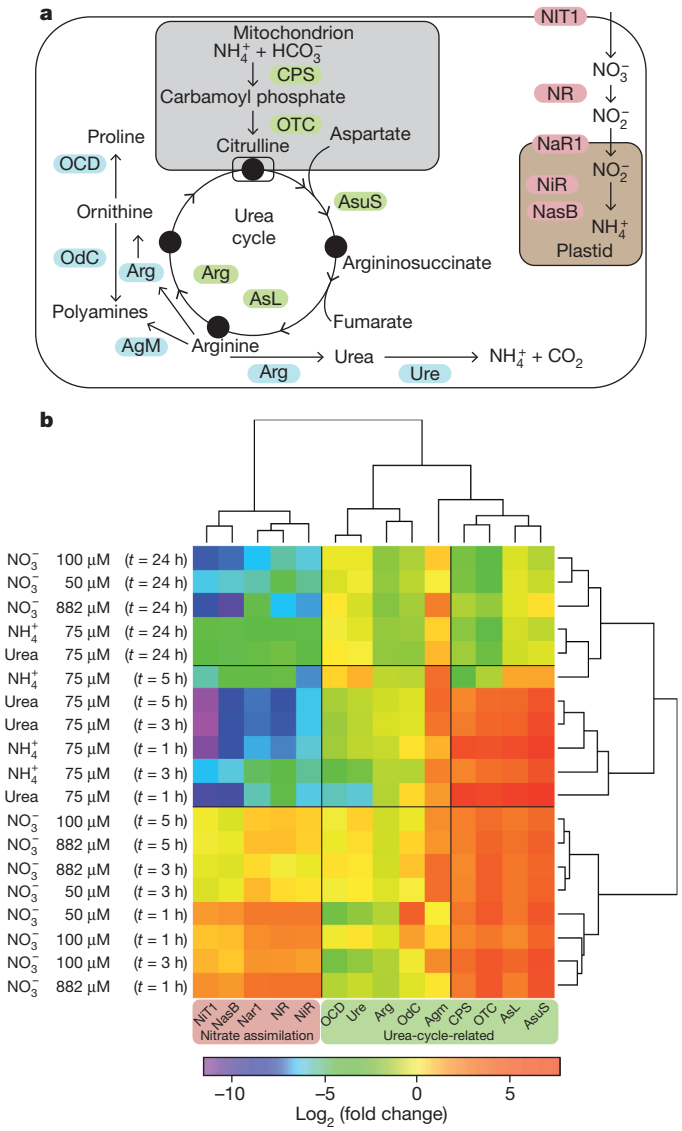

Figure 2 | Transcript levels of genes encoding components of the urea cycle and the nitrate assimilation pathway in wild-type *P. tricornutum* cells. **a**, Schematic showing the roles of urea-cycle and nitrate-assimilation enzymes. Agm, agmatinase; Arg, arginase; AsL, argininosuccinate lyase; AsuS, argininosuccinate synthase; CPS, carbamoyl phosphate synthase; Nar1, plastid nitrite transporter; NasB, NADPH nitrite reductase; NiR, ferredoxin nitrite reductase; NiT1, nitrate transporter; NR, nitrate reductase; OCD, ornithine cyclodeaminase; OdC, ornithine decarboxylase; OTC, ornithine transcarboxylase; Ure, urease. (Arginase is depicted twice because the arginase reaction results in ornithine and urea.) **b**, Hierarchical clustering of transcriptional fold changes, from triplicate technical replicates of duplicate cultures ($n = 2$), in response to the addition of nitrogen substrates to nitrogen-limited cultures over a 12 h:12 h light:dark cycle. Log_2 values $> (\pm) 2$ (fourfold) are significant ($P < 0.05$).

influence protein turnover by facilitating the reallocation of inorganic carbon and nitrogen from protein catabolism.

Subsequent experiments examined the abundance of numerous metabolites during recovery from nitrogen deprivation (Fig. 3a–d). Metabolite profiles for nitrogen-deprived RNAi lines relative to wild-type cells 6 h after exposure to nitrogen-replete media showed significant lags in the cellular accumulation of most metabolites detected, normalized to total protein (Supplementary Fig. 10). In wild-type cells, the end of exponential growth is associated with the maximum accumulation of cellular metabolites, particularly tricarboxylic acid cycle (TCA) intermediates and the OUC derivatives ornithine, urea, spermidine and proline (Supplementary Fig. 11). Relative to wild-type cultures, the RNAi lines exhibited an initial lag in metabolite accumulation, followed by a steady increase and then a subsequent crash (Fig. 3d and Supplementary Fig. 12). Five of the six most severely affected metabolites are amino acids with direct connections to the OUC: glutamine, asparagine, aspartic acid, arginine and ornithine. A reduction in nitrogen input to the OUC should affect arginine and urea most directly because nitrogen atoms entering the OUC subsequently reside in arginine and ultimately, in urea. Arginine, in turn, is the precursor for ornithine and proline, which together accounted for more than 70% of the nitrogen and carbon contained in metabolites detected in wild-type cells (Supplementary Fig. 13a, b). Proline is known to be the principal osmolyte in diatoms[12] and arginine and ornithine are precursors for important cell-wall-associated long-chain polyamine components such as spermidine[11], which is also significantly affected by unCPS deficiency (Fig. 3d and Supplementary Fig. 10). Many metabolic offshoots of the OUC are unique to diatoms, being enabled by genes acquired by lateral transfer from bacteria[6,7,9], such as the genes encoding ornithine cyclodeaminase, agmatinase, glutamine synthetase III, carbamate kinase and hydroxylamine reductase. These genes have significantly expanded the functionality of the diatom OUC.

In animal cells, an association between the OUC and TCA cycles through the aspartate–argininosuccinate shunt is well established[26] and its extension to photosynthetic microalgae is intriguing. Correlation analyses of TCA and OUC metabolites in wild-type cells indicated that the OUC derivatives urea and proline are tightly coupled to TCA cycle intermediates; except for the intermediate 2-oxoglutarate, which is associated with glutamine (Supplementary Fig. 14). In the *unCPS* knockdown lines, however, TCA cycle intermediates were completely uncoupled from the OUC endpoints, urea and proline (Supplementary Fig. 14). The interdependence between the OUC and TCA cycles in wild-type *P. tricornutum* cells is exemplified by the highly similar impacts of mitochondrial *unCPS* knockdown on OUC-related metabolites and on the TCA cycle intermediates fumarate and malate, which are downstream of the aspartate–argininosuccinate shunt. The impact of this knockdown upon TCA cycle metabolites that occur upstream of the aspartate–argininosuccinate shunt, such as citric acid and succinic acid, is not as severe (Fig. 3d), perhaps owing to a reduction in fumarate regeneration. The significant accumulation of 2-oxoglutarate in the RNAi lines relative to the wild-type is consistent with a severe shortage of the critical nitrogen metabolite glutamine (five to sixfold reduction in RNAi lines)[27]. Taken together, these data illustrate important connections between the OUC, the glutamine synthetase/glutamate synthase cycle and the TCA cycle in diatoms (Fig. 4).

Coupling between the OUC and TCA cycles may be due to a shared source of precursor metabolites. Mitochondrial amino acid catabolism yields carbon skeletons for the TCA cycle as well as NH_4^+ and HCO_3^- that may be scavenged by mitochondrial CPS. Photorespiration also generates NH_4^+ and HCO_3^- in mitochondria[6,27,28]. Our results implicate the diatom OUC as an important cellular hub for redistribution and turnover of inorganic carbon and nitrogen from catabolism and/or photorespiration. Proline and glutamine, which are critical metabolites for diatom cell wall formation, chitin synthesis and overall cellular nitrogen status in photosynthetic cells[11,27], are especially affected by CPS impairment in diatoms.

The ability to link catabolic and anabolic metabolism efficiently through turnover and reallocation of intracellular carbon and nitrogen into key cellular components such as polyamines and glutamine, which enhance diatom growth and defence, is inherently crucial for the rapid proliferation of diatoms after upwelling events. Our results show that the diatom OUC is an important anaplerotic carbon-fixation pathway that facilitates this process. The pattern of *CPS* duplication and evolution indicates that core OUC biochemistry probably evolved in the exosymbiont before plastid acquisition, thereby excluding it from the Plantae (Archaeplastida) lineage. The integration of the OUC into core carbon and nitrogen metabolism and its connection to the synthesis of

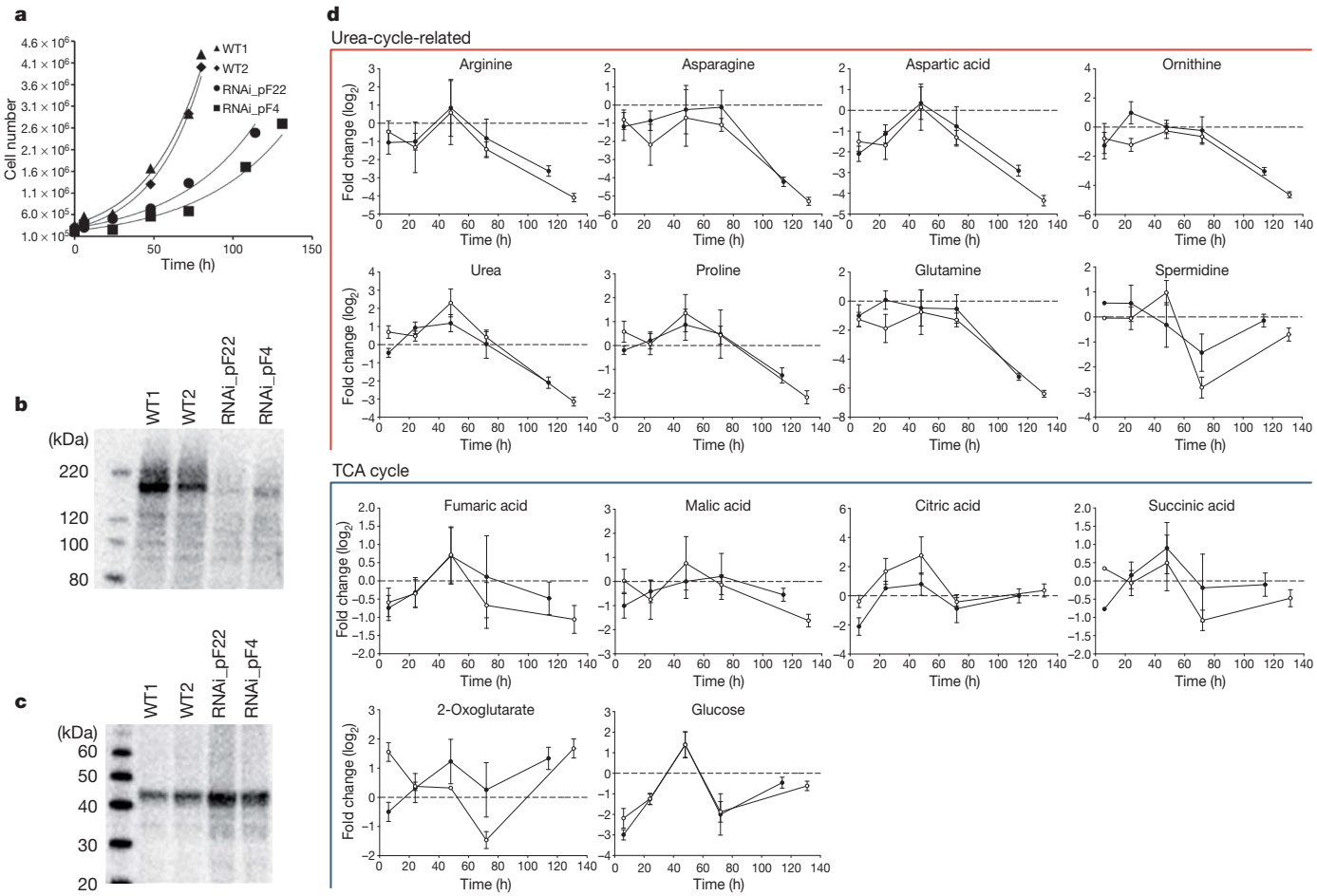

Figure 3 | Growth characteristics and metabolite abundance in wild-type and *unCPS* RNAi *P. tricornutum* lines. **a**, Growth curves of wild-type (WT1, WT2) and RNAi (pF22, pF4) lines. **b**, **c**, Western blot analyses of unCPS (**b**) and alternative oxidase (**c**) in wild-type and RNAi lines at $t = 24$ h. **d**, Fold changes in metabolites related to the urea cycle and citric acid cycle in two *unCPS* RNAi lines relative to wild type. Wild-type metabolite levels were calculated as the mean of triplicate technical replicates performed on duplicate cultures ($n = 2$). Fold changes result from the comparison of metabolite levels at time points $t = 6$ h, 24 h, 48 h, 72 h and 80 h in wild-type lines with metabolites in the RNAi lines at time points $t = 6$ h, 24 h, 72 h, 114 h and 131 h. Error bars, s.d.

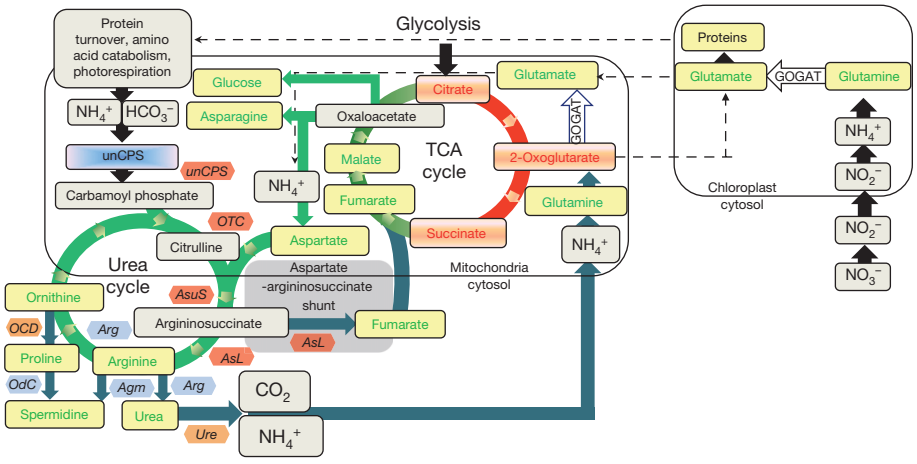

Figure 4 | Conceptual overview of the roles of unCPS and the diatom urea cycle on the basis of metabolite data from wild-type and RNAi lines. Metabolites depicted in green were significantly depleted (*t*-test, $P < 0.05$) whereas those shown in red were less affected in *unCPS* RNAi lines. Green and blue arrows indicate fluxes hypothesized to be particularly strongly affected by diatom unCPS impairment. Blue arrows further indicate potentially critical control points that link the urea cycle to other major metabolic pathways. Core urea cycle genes that displayed coordinated expression, *AsL*, *AsuS*, *OTC* and *unCPS*, are indicated in red hexagons. *OCD* and *Ure*, depicted in orange hexagons, as well as *Arg*, *OdC* and to some extent *Agm*, depicted in blue hexagons, also showed independent coordination in overall expression. GOGAT, glutamate synthase.

key cellular metabolites may provide a partial explanation for the dominance of diatoms in the modern ocean.

METHODS SUMMARY

Cultures were grown in F/2 seawater with modified nitrogen substrates and levels. A full-length GST–CPS fusion product was purified from soluble *Escherichia coli* protein lysates using the MagneGST protein purification system (Promega). CPS activity was evaluated by monitoring the rate of conversion of [^{14}C]$NaHCO_3$ to [^{14}C]carbamoyl phosphate[13]. *CPS* RNAi lines were generated by construction of RNAi expression vectors with overlapping inverted repeat fragments (Supplementary Fig. 15) of the *P. tricornutum CPS* gene that were unique in the genome[24]. Western blot analyses were performed with custom *P. tricornutum* genomic antibodies (Strategic Diagnostics) for CPS, alternative oxidase, nitrate reductase and urease. For quantitative reverse transcription PCR (qRT–PCR) experiments, 100–500 ng total RNA was reverse-transcribed and amplicons were quantified in qRT–PCR reactions. All PCR primers are given in Supplementary Table 2. Metabolites were extracted with methanol, derivatized and analysed by gas chromatography mass spectrometry[29] Maximum likelihood phylogeny was inferred from CPS amino acid sequences, computed using RAxML 7.2.6[30] under the LG + Gamma model of evolution. See Methods for details on estimates of CPS divergence times.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 4 December 2010; accepted 24 March 2011.

1. Falkowski, P. G. *et al.* The evolution of modern eukaryotic phytoplankton. *Science* **305,** 354–360 (2004).
2. Falkowski, P. G. & Oliver, M. J. Mix and match: how climate selects phytoplankton. *Nature Rev. Microbiol.* **5,** 813–819 (2007).
3. Nelson, D. M., Treguer, P., Brzezinski, M. A., Leynaert, A. & Queguiner, B. Production and dissolution of biogenic silica in the ocean - revised global estimates, comparison with regional data and relationship to biogenic sedimentation. *Glob. Biogeochem. Cycles* **9,** 359–372 (1995).
4. Smetacek, V. Diatoms and the ocean carbon cycle. *Protist* **150,** 25–32 (1999).
5. Hamm, C. & Smetacek, V. in *Evolution of Primary Producers in the Sea* (eds Falkowski, P. G. & Knoll, A. H.) (Academic Press, 2007).
6. Allen, A. E., Vardi, A. & Bowler, C. An ecological and evolutionary context for integrated nitrogen metabolism and related signaling pathways in marine diatoms. *Curr. Opin. Plant Biol.* **9,** 264–273 (2006).
7. Bowler, C. *et al.* The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456,** 239–244 (2008).
8. Moustafa, A. *et al.* Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science* **324,** 1724–1726 (2009).
9. Bowler, C., Vardi, A. & Allen, A. E. Oceanographic and biogeochemical insights from diatom genomes. *Ann. Rev. Mar. Sci.* **2,** 333–365 (2010).
10. Allen, A. E. *et al.* Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation. *Proc. Natl Acad. Sci. USA* **105,** 10438–10443 (2008).
11. Kröger, N. & Poulsen, N. Diatoms—from cell wall biogenesis to nanotechnology. *Annu. Rev. Genet.* **42,** 83–107 (2008).
12. Krell, A., Funck, D., Plettner, I., John, U. & Dieckmann, G. Regulation of proline metabolism under salt stress in the psychrophilic diatom *Fragilariopsis cylindrus* (Bacillariophyceae). *J. Phycol.* **43,** 753–762 (2007).
13. Anderson, P. M. Glutamine- and N-acetylglutamate-dependent carbamoyl phosphate synthetase in elasmobranchs. *Science* **208,** 291–293 (1980).
14. Hong, J., Salo, W. L., Lusty, C. J. & Anderson, P. M. Carbamoyl-phosphate synthetase-III, an evolutionary intermediate in the transition between glutamine-dependent and ammonia-dependent carbamoyl-phosphate synthetases. *J. Mol. Biol.* **243,** 131–140 (1994).
15. Mommsen, T. P. & Walsh, P. J. Evolution of urea synthesis in vertebrates: the piscine connection. *Science* **243,** 72–75 (1989).
16. Lawson, F. S., Charlebois, R. L. & Dillon, J. A. R. Phylogenetic analysis of carbamoylphosphate synthetase genes: Complex evolutionary history includes an internal duplication within a gene which can root the tree of life. *Mol. Biol. Evol.* **13,** 970–977 (1996).
17. Guppy, M. The hibernating bear: why is it so hot, and why does it cycle urea through the gut. *Trends Biochem. Sci.* **11,** 274–276 (1986).
18. Holden, H. M., Thoden, J. B. & Raushel, F. M. Carbamoyl phosphate synthetase: an amazing biochemical odyssey from substrate to product. *Cell. Mol. Life Sci.* **56,** 507–522 (1999).
19. Armbrust, E. V. *et al.* The genome of the diatom *Thalassiosira pseudonana*: Ecology, evolution, and metabolism. *Science* **306,** 79–86 (2004).
20. Ast, M. *et al.* Diatom plastids depend on nucleotide import from the cytosol. *Proc. Natl Acad. Sci. USA* **106,** 3621–3626 (2009).
21. Maheswari, U. *et al.* Digital expression profiling of novel diatom transcripts provides insight into their biological functions. *Genome Biol.* **11,** R85 (2010).
22. Esteban-Pretel, G. *et al.* Vitamin A deficiency increases protein catabolism and induces urea cycle enzymes in rats. *J. Nutr.* **140,** 792–798 (2010).
23. Lee, B. *et al. In vivo* urea cycle flux distinguishes and correlates with phenotypic severity in disorders of the urea cycle. *Proc. Natl Acad. Sci. USA* **97,** 8021–8026 (2000).
24. De Riso, V. *et al.* Gene silencing in the marine diatom *Phaeodactylum tricornutum*. *Nucleic Acids Res.* **37,** e96 (2009).
25. Young, E. B. & Beardall, J. Photosynthetic function in *Dunaliella tertiolecta* (Chlorophyta) during a nitrogen starvation recovery cycle. *J. Phycol.* **39,** 897–905 (2003).
26. Morris, S. M. Regulation of enzymes of the urea cycle and arginine metabolism. *Annu. Rev. Nutr.* **22,** 87–105 (2002).
27. Nunes-Nesi, A., Fernie, A. R. & Stitt, M. Metabolic and signaling aspects underpinning the regulation of plant carbon nitrogen interactions. *Mol. Plant* **3,** 973–96 (2010).
28. Parker, M. S., Mock, T. & Armbrust, E. V. Genomic insights into marine microalgae. *Annu. Rev. Genet.* **42,** 619–645 (2008).
29. Lisec, J., Schauer, N., Kopka, J., Willmitzer, L. & Fernie, A. R. Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nature Protocols* **1,** 387–396 (2006).
30. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and models. *Bioinformatics* **22,** 2688–2690 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Meichenin and C. Lichtlé for assistance with electron microscropy, J. C. Thomas for CPS purification and activity experiments, A. Falciatore for advice on RNAi constructs and A. Main for screening and evaluation of RNAi lines. This study was supported by the National Science Foundation (NSF-OCE-0722374, NSF-OCE-0727997, NSF-MCB-1024913) and JCVI internal funding to A.E.A., the European Commission Diatomics project and Agence Nationale de la Recherche (France) (C.B.) and the Czech Science Foundation (206/08/1423) (M.O.).

Author Contributions A.E.A. and C.B. designed the study. A.E.A. performed CPS localization, confocal microscopy, protein purification, activity, overexpression and other laboratory experiments. A.E.A. and C.L.D. designed nitrogen-recovery experiments and the physiological characterization of the RNAi as well as wild-type experiments, which were performed by H.Z and C.L.D.. H.Z. generated and screened RNAi lines. H.Z. and H.H. performed long-term and short-term nitrogen-recovery and related qPCR experiments. D.A.J. ran qPCR reactions and assisted with analyses of qPCR data. M.O., A.H. and A.E.A. generated and analysed phylogenetic and molecular clock data. A.N-N. and A.R.F. performed metabolite profiling of samples collected from RNAi and wild-type cultures. J.P.M., C.L.D. and A.E.A. analysed qPCR, metabolite and western blot data in detail. A.E.A. wrote the paper with assistance from C.L.D., A.R.F., C.B. and M.O. All the authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.E.A. (aallen@jcvi.org).

METHODS

Culturing. All cultures were grown in F/2 media with modified nitrogen substrates and levels. Short-term qRT–PCR experiments were performed on cultures grown under 12 h:12 h light:dark cycles and all other experiments were conducted in constant light. For short-term qRT–PCR experiments (Fig. 2), cells from an exponential culture grown in standard F/2 media (882 μM NO_3^-) were collected, washed twice with nitrogen-free media and then inoculated at 4×10^{-5} cells ml^{-1} in media containing 50 μM NO_3^-. Cell number was monitored with cell counts and after 10 days at stationary phase, nitrogen-starved cells were collected and washed twice with nitrogen-free media, then duplicate cultures were inoculated into media containing different nitrogen sources (50, 100 or 882 μM NO_3^-, 75 μM NH_4^+ or 75 μM urea). Through a diel cycle, cells were collected for RNA extractions at $t = 0$ (the beginning of the 12 h:12 h light period), then at 3 h, 5 h and 24 h.

For longer-term nitrogen recovery experiments, quantitative PCR (qPCR) and western blot experiments (Supplementary Fig. 4), pre-culturing was in media containing 100 μM NH_4^+ as the sole nitrogen source. PO_4^{-3} was adjusted to 37.5 μM to force cells into nitrogen limitation owing to a low nitrogen:phosphorus ratio. Cultures were maintained in HEPES-buffered media (1 mM, pH 8.1), agitated with internal stir bars and bubbled with air. After NH_4^+ concentrations declined to below detection (500 nM), the cultures reached stationary phase. After n9 days in stationary phase, Fv/Fm levels declined to approximately 0.1. Duplicate cultures then received additions of either 500 μM NO_3^-, 500 μM NH_4^+ or 250 μM urea. RNA and protein samples were collected at $t = 0$ h, 6 h, 24 h, 28 h and 120 h.

For all other nitrogen-recovery experiments, culture flasks were maintained on stir plates, bubbled with air, and pre-culturing was in media with urea as the sole nitrogen source. As above, urea was added to a final concentration of 50 μM, resulting in a media nitrogen:phosphorus ratio of 10:3.7. After a period of time in stationary phase that varied according to the experiment, cells were collected, rinsed with nitrogen-free media and inoculated into media containing 880 μM NO_3^- as the sole nitrogen source. For physiological experiments designed to measure the impact of CPS impairment on the capacity to recover from nitrogen limitation (Supplementary Figs 8 and 9), the period of stationary phase for the pre-cultures ranged from 7 to 13 days. For metabolite-profiling experiments (Fig. 3 and Supplementary Figs 10–14), the duration of the light period for stationary phase pre-cultures was reduced to 3 days so that no discernable lag phase was apparent between RNAi and wild-type cultures. To sample comparable points in the growth curve, RNAi and wild-type cells in these experiments were exposed to a reduced period of stationary phase pre-culturing and the duration of the lag phase before the onset of exponential growth, in terms of cell number and fluorescence, was negligible. According to Fv/Fm, cell number and fluorescence, all of the time points sampled for all cell lines in these experiments were during exponential growth.

Steady growth rates for wild-type and RNAi lines were obtained from growth in media with either nitrate or urea as the sole nitrogen source. Acclimated specific growth rate (μ) values were estimated from *in vivo* chlorophyll-*a* fluorescence (10 AU, Turner Instruments). Maximum photochemical yields of photosystem II were determined with a Water-PAM fluorometer (Walz). Cell counts were performed by microscopy using a Sedgwich-Rafter slide or a Beckman Quanta flow cytometer.

Cloning, RNAi specificity and transgenic expression. *Phaeodactylum tricornutum* cDNA encoding full-length CPS was PCR-amplified and cloned into a TOPO pENTR vector (Invitrogen). A clone containing an error-free sequence was selected for Gateway recombination (Invitrogen) with a diatom C-terminal YFP pDONR vector[31]. The resulting expression vector was transformed into *P. tricronutum* by particle bombardment[32]. Transformants were then screened by PCR and used for confocal imaging and electron microscopy immunolocalization using an antiserum against GFP. CPS pENTR vectors were also recombined with a pDONR vector designed for N-terminal GST tagging and expression in BL21A *Escherichia coli* cells. *CPS* RNAi lines were generated by construction of RNAi expression vectors with overlapping inverted repeat fragments (Supplementary Fig. 15) of the *P. tricornutum CPS* gene, followed by transformation in *P. tricornutum*[24]. The 457-base-pair gene fragment used in the RNAi expression cassette is specific for *P. tricornutum unCPS* and is not homologous to any other region of the *P. tricronutum* genome. All cloning primer sets are listed in Supplementary Table 2.

CPS activity. A full-length GST–CPS fusion product was purified from soluble *E. coli* protein lysates using the MagneGST protein purification system (Promega). CPS activity was evaluated by monitoring the rate of conversion of [^{14}C]$NaHCO_3$ to [^{14}C]carbamoyl phosphate. [^{14}C]carbamoyl phosphate formed after 30 min at 20 °C was measured after conversion to [^{14}C] urea by boiling and addition of NH_4Cl, followed by removal of [^{14}C]$NaHCO_3$ and other ^{14}C-labelled anions by elution from an anion exchange resin[13].

RNAi screening, protein levels, gene expression and metabolites. Western blot analyses were performed with custom *P. tricornutum* Genomic Antibodies (Strategic Diagnostics) for CPS, nitrate reductase, urease and alternative oxidase. Screening of RNAi lines for protein levels was performed on lines which initially showed a reduction in growth on urea relative to wild-type. Subsequently, these lines were evaluated for CPS and alternative oxidase protein levels. Six lines that showed CPS deficiency did not show depleted alternative oxidase levels. For qPCR experiments, 100–500 ng total RNA was reverse transcribed using the Quantitect Reverse Transcription Kit (Qiagen) and amplicons were quantified in qRT–PCR reactions with Fast SYBR Green Master Mix (Applied Biosystems). qRT–PCR primer sets were evaluated for optimal efficiency and the $2^{-\Delta\Delta CT}$ method was used to estimate fold change in gene expression, normalized to three endogenous control genes, 18S rDNA, TATA binding protein and histone H4[31]. ΔCT values were obtained by subtracting the mean values of experimental genes from a mean of the control genes for each sample. Using a Welch approximation for unequal group variances, a *P* value was estimated on the basis of the *t*-distribution that resulted from a between-subjects *t*-test evaluating the control RNA ($t = 0$) relative to a given experimental RNA. qRT–PCR primers are listed in Supplementary Table 2. Metabolites were extracted with methanol, derivatized and analysed by gas chromatography mass spectrometry[29]. Metabolites were identified by comparison with database entries of authentic mass-spectral and retention-index (MSRI) libraries[33]. Hierarchial clustering of gene expression and metabolite data are according to the $\log_2$ of gene expression fold changes and metabolite correlation coefficients, which were plotted using the 'heatmap' function in R[34] without additional scaling. Hierarchical clustering of the expression fold changes and metabolite correlation as level vectors are shown as dendrograms and are the basis for the ordering of experimental conditions (columns) and genes (rows).

Phylogenetic analysis. CPS amino acid sequences (Supplementary Table 5) were aligned using Mafft 6[35]. The alignment was edited in Seaview 4[36]. We tested all sequences for deviation from average amino acid composition using a chi-squared-based homogeneity test as implemented in Tree-Puzzle 5.2[37]. Because some taxa that were important for interpretation did not pass the homogeneity test (see Supplementary Table 3 for a list), we decided to perform phylogenetic analyses using both empirical and mixture models. The maximum likelihood (ML) topology was computed using RAxML 7.2.6[30] under the LG + Gamma model of evolution. The tree with the highest likelihood score was chosen from 200 independent analyses starting with randomized parsimony trees. Non-parametric bootstrap support was estimated from 500 iterations. In addition, ML analysis was performed using an empirical profile mixture model (C40) as implemented in PhyML-CAT[38–40]. The topology with the highest likelihood score from 15 independent runs starting with randomized trees using both NNI and SPR swapping algorithms was selected. Owing to the considerable computational burden, we performed an approximate likelihood-ratio test on the 'best' tree (that is, aLRT)[41] instead of more traditional non-parametric bootstrap analysis. We also analysed our dataset using a Bayesian approach; again, with both empirical and mixture models. For the empirical approach, we used MrBayes 3.1[42,43] with two chains run under the WAG + Gamma model for two million generations; priors were set to defaults and the first 500,000 generations were omitted from tree reconstruction as a burn-in. Four chains were run until they reached convergence in Phylobayes 3.2d[44] to infer the topology using the mixture model combined with LG empirical exchange rates (CAT-LG model).

Divergence time estimation. The fit of the clock-like behaviour to the 'best' tree was tested and rejected by likelihood-ratio test as implemented in Tree-Puzzle 5.2. We therefore used algorithms that enable relaxation of the molecular clock to infer the date of duplication events from the phylogenetic tree. The penalized likelihood (PL) and non-parametric-smoothing (NPRS) algorithms implemented in R8S 1.71[45,46], as well as a log-normal auto-correlated relaxed clock model[47] implemented in Phylobayes 3.2, were used to estimate divergence times. Because the PL and NPRS yielded virtually the same results, we show only those of PL. The fossil calibration points compiled from[48,49] are listed in Supplementary Table 4.

31. Siaut, M. *et al.* Molecular toolbox for studying diatom biology in *Phaeodactylum tricornutum*. *Gene* **406,** 23–35 (2007).
32. Falciatore, A., Casotti, R., Leblanc, C., Abrescia, C. & Bowler, C. Transformation of nonselectable reporter genes in marine diatoms. *Mar. Biotechnol.* **1,** 239–251 (1999).
33. Schauer, N. *et al.* GC-MS libraries for the rapid identification of metabolites in complex biological samples. *FEBS Lett.* **579,** 1332–1337 (2005).
34. R Development Core Team. R: A language and environment for statistical computing. ISBN 3-900051-07-0 (R Foundation for Statistical Computing, 2008).
35. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30,** 3059–3066 (2002).
36. Gouy, M., Guindon, S. & Gascuel, O. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27,** 221–224 (2010).
37. Schmidt, H. A., Strimmer, K., Vingron, M. & von Haeseler, A. TREE-PUZZLE: a maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18,** 502–504 (2002).
38. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52,** 696–704 (2003).
39. Lartillot, N. & Philippe, H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* **21,** 1095–1109 (2004).

40. Quang, L. S., Gascuel, O. & Lartillot, N. Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics* **24,** 2317–2323 (2008).
41. Anisimova, M. & Gascuel, O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst. Biol.* **55,** 539–552 (2006).
42. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* **17,** 754–755 (2001).
43. Ronquist, F. & Huelsenbeck, J. P. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19,** 1572–1574 (2003).
44. Lartillot, N., Lepgae, T. & Blanquart, S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25,** 2286–2288 (2009).
45. Sanderson, M. J. A nonparametric approach to estimating divergence times in the absence of rate consistency. *Mol. Biol. Evol.* **19,** 1218–1231 (1997).
46. Sanderson, M. J. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol. Biol. Evol.* **19,** 101–109 (2002).
47. Thorne, J. L., Kishino, H. & Painter, I. S. Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.* **15,** 1647–1657 (1998).
48. Berney, C. & Pawlowski, J. A molecular timescale for eukaryote evolution recalibrated with the continuous microfossil record. *Proc. R. Soc. A/B* **273,** 1867–1872 (2006).
49. Benton, M. J. & Donoghue, P. C. J. Paleontological evidence to date the tree of life. *Mol. Biol. Evol.* **24,** 26–53 (2007).